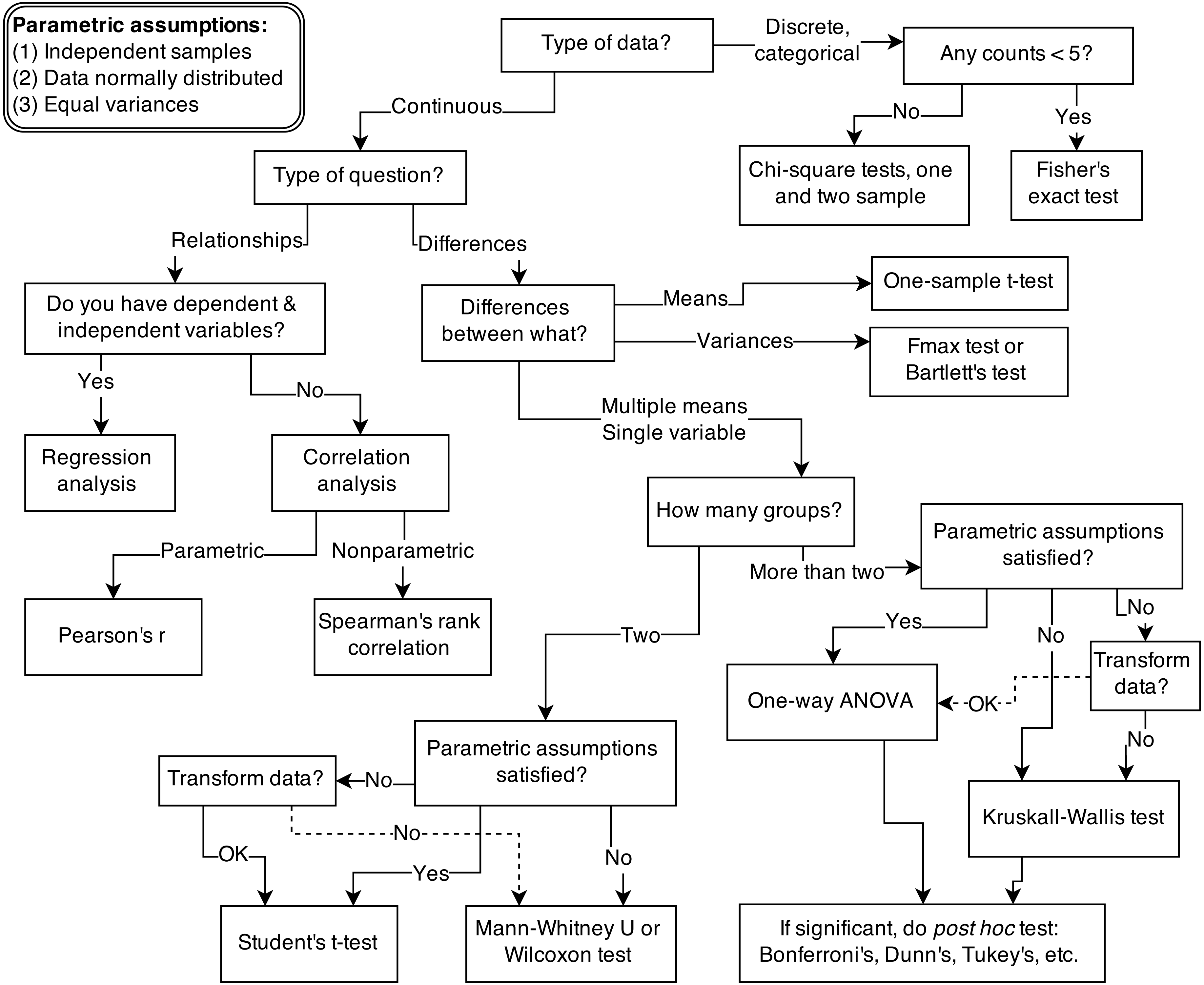


Parametric assumptions:
(1) Independent samples
(2) Data normally distributed
(3) Equal variances
Type of data?
Discrete, categorical
Any counts < 5?
No
Yes
Chi-square tests, one and two sample
Fisher's exact test
Continuous
Type of question?
Relationships
Differences
Do you have dependent & independent variables?
Yes
No
Regression analysis
Correlation analysis
Parametric
Nonparametric
Pearson's r
Spearman's rank correlation
Differences between what?
Means
One-sample t-test
Variances
Fmax test or Bartlett's test
Multiple means Single variable
How many groups?
More than two
Parametric assumptions satisfied?
Yes
No
No
Transform data?
OK
No
One-way ANOVA
Kruskall-Wallis test
If significant, do post hoc test: Bonferroni's, Dunn's, Tukey's, etc.
Two
Parametric assumptions satisfied?
No
Transform data?
OK
No
Yes
No
Student's t-test
Mann-Whitney U or Wilcoxon test

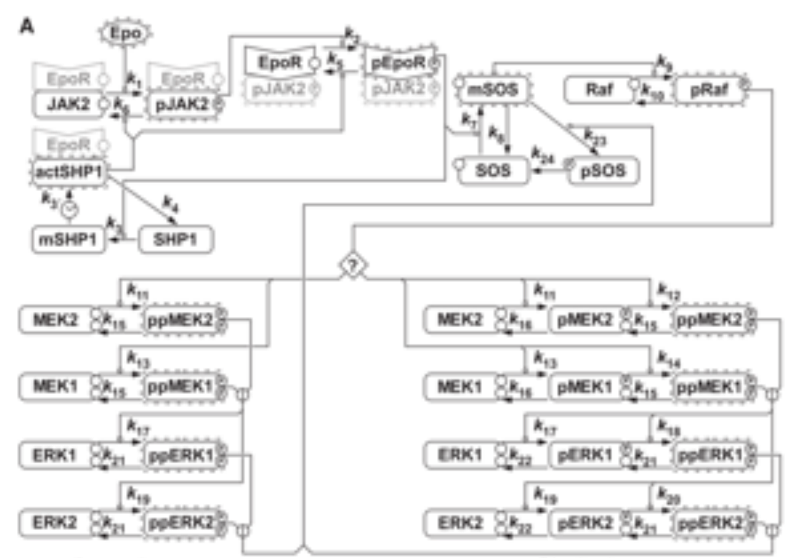
The Golems of Science

Golem

- Made of clay
- Animated by “truth”
- Powerful
- Blind to creator’s intent
- Easy to misuse
- Fictional

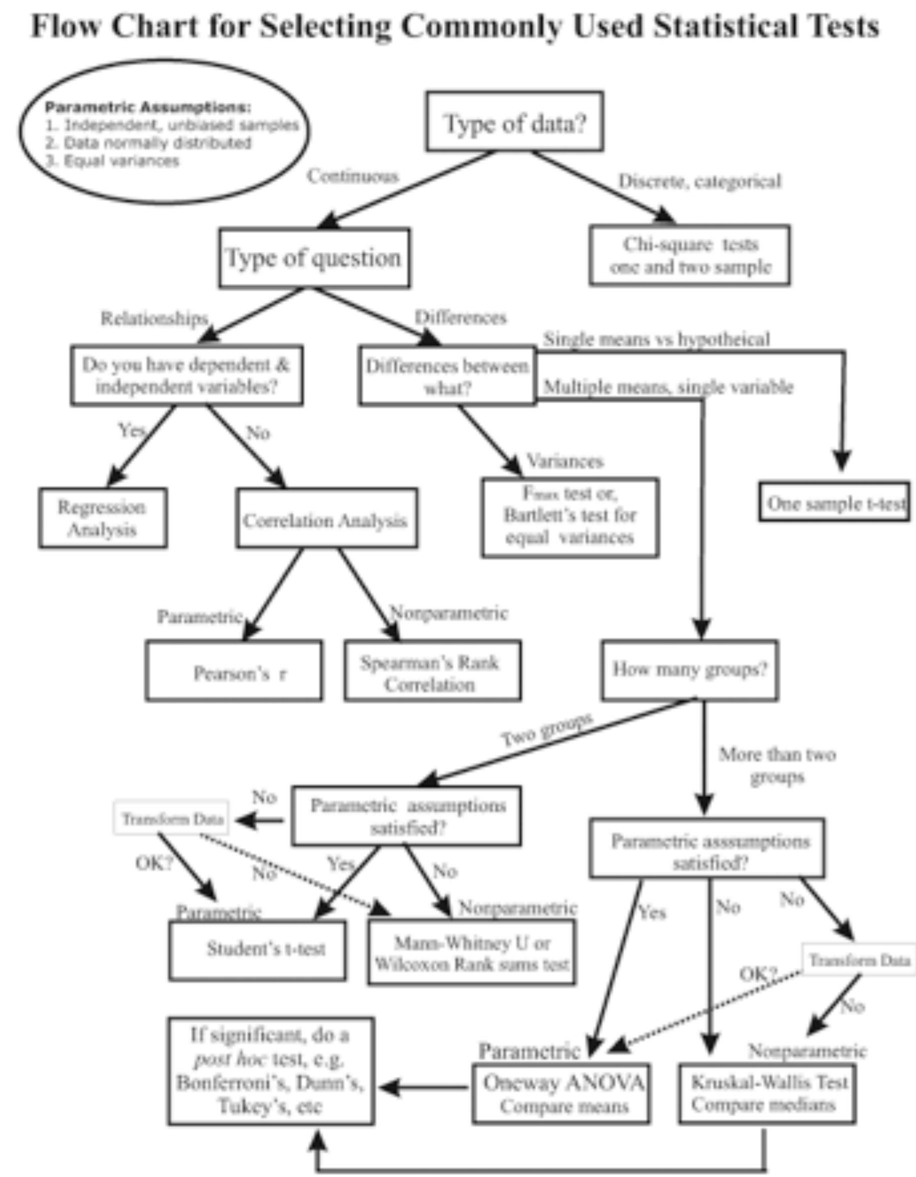
Model

- Made of...silicon?
- Animated by “truth”
- Hopefully powerful
- Blind to creator’s intent
- Easy to misuse
- Not even false



Against Tests

- Specialized, pre-made golems, “procedures”
- Most developed in early 20th century, fragile, eclipsed by more recent tools
- Users don’t know they are using models (golems)
- Falsifying *null* model not sufficient
- Inference is not decision



O, that way madness lies

Golem Engineering

- Need a framework for developing and vetting statistical golems
- Several options
- We'll use this one
  - Bayesian data analysis
  - Multilevel modeling
  - Model comparison

From *Breath of Bones: A Tale of the Golem*

Bayesian data analysis

- Use *probability* to describe uncertainty
  - Extends ordinary logic (true/false) to continuous *plausibility*
- Computationally difficult
  - Markov chain Monte Carlo (MCMC) to the rescue
- Used to be controversial
  - Ronald Fisher: Bayesian analysis "must be wholly rejected."

Pierre-Simon Laplace (1749–1827)

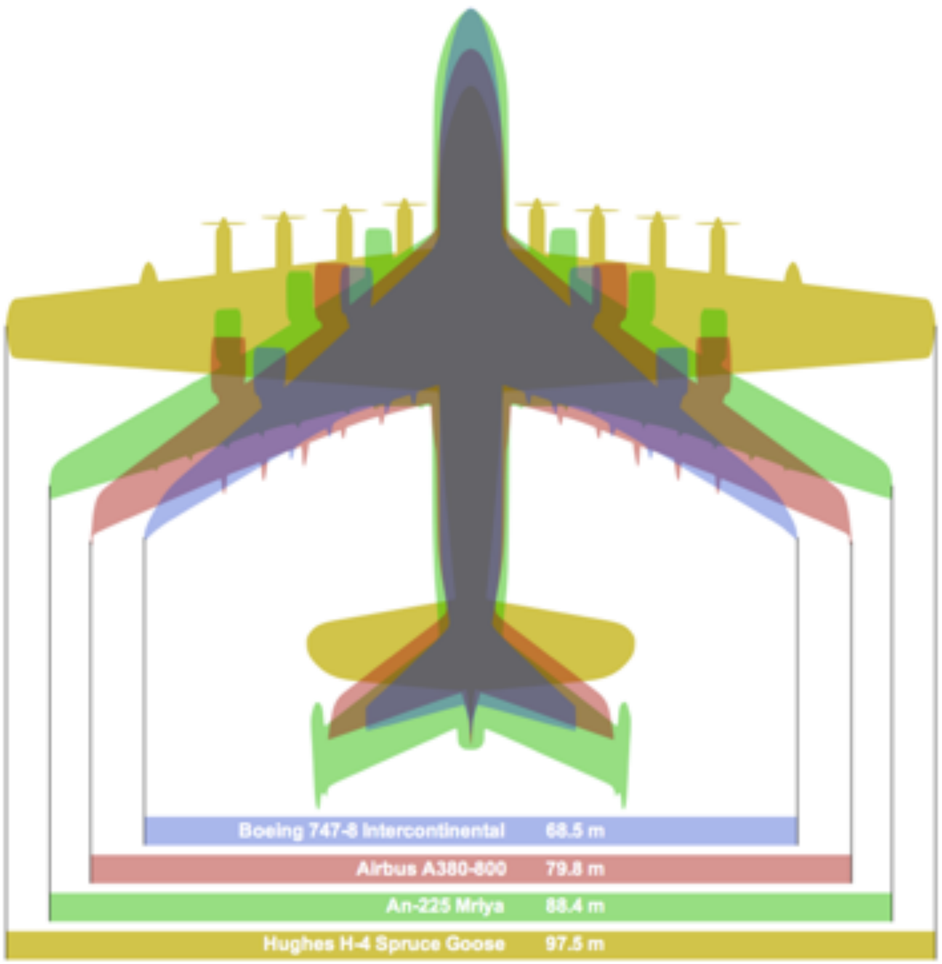
Sir Harold Jeffreys (1891–1989) with Bertha Swirles, aka Lady Jeffreys (1903–1999)

Multilevel models

- Models with *multiple levels* of uncertainty
  - Replace parameters with models
- Common uses
  - Repeat & imbalanced sampling
  - Study variation
  - Avoid averaging
  - Phylogenetics, factor and path analysis, networks, spatial models
- Natural Bayesian strategy

Model comparison

- Instead of falsifying a **null** model, compare **meaningful** models
- Basic problems
  - Overfitting
  - Causal inference
- Ockham's razor is silly
- Information theory less silly
  - AIC, WAIC, cross-validation
- Must distinguish prediction from inference



Bayesian data analysis

Count all the ways data can happen, according to assumptions.

Assumptions with more ways that are consistent with data are more plausible.

Garden of Forking Data

- The future:
  - Full of branching paths
  - Each choice closes some
- The data:
  - Many possible events
  - Each observation eliminates some

Building a model

- How to use probability to do typical statistical modeling?

1. Design the model (data story)
2. Condition on the data (update)
3. Evaluate the model (critique)

Nine tosses of the globe:

W L W W W L W L W

Design > Condition > Evaluate

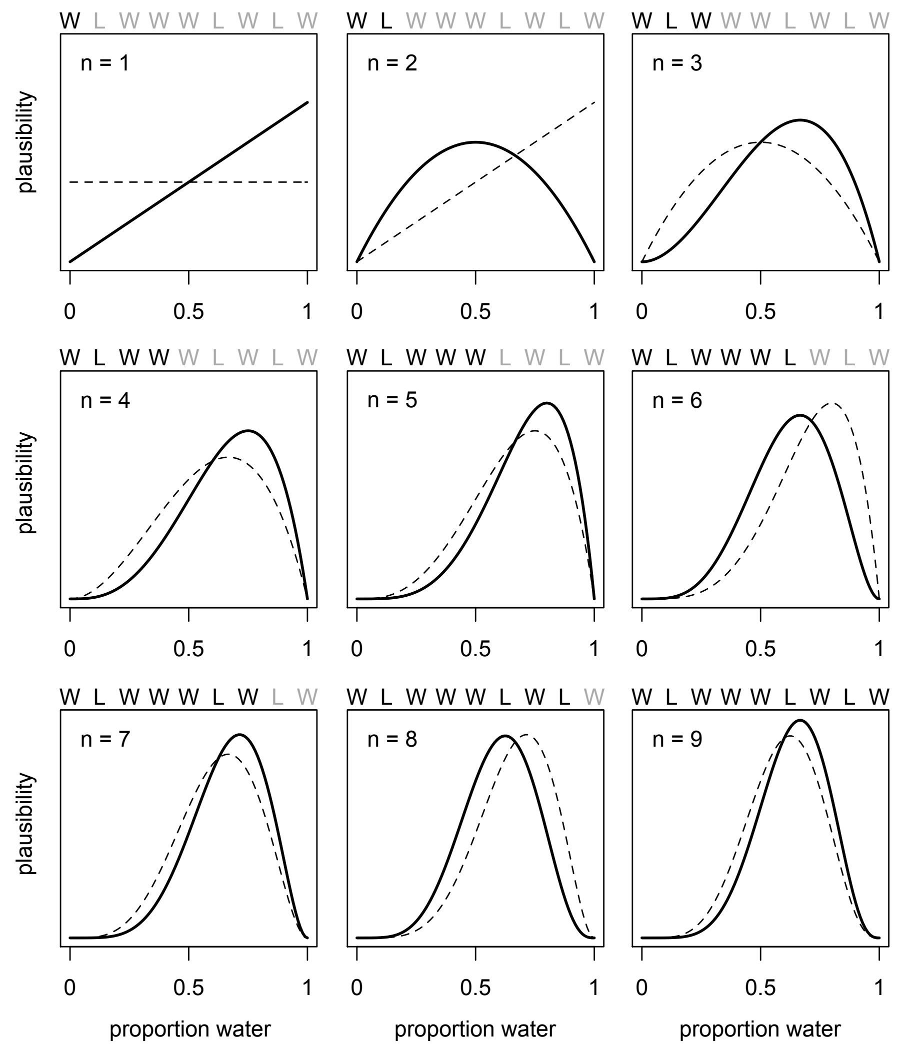
- Data story motivates the model
  - How do the data arise?
- For **W L W W W L W L W**:
  - Some true proportion of water, p
  - Toss globe, probability p of observing W, $1-p$ of L
  - Each toss therefore independent of other tosses
- Translate data story into probability statements

Design > Condition > Evaluate

- *Bayesian updating* defines optimal learning in small world, converts *prior* into *posterior*
  - Give your golem an information state, before the data: Here, an initial confidence in each possible value of *p* between zero and one
  - Condition on data to update information state: New confidence in each value of *p*, conditional on data

Design > Condition > Evaluate

- Data order irrelevant, because golem assumes order irrelevant
  - All-at-once, one-at-a-time, shuffled order all give same posterior
- Every posterior is a prior for next observation
- Every prior is posterior of some other inference
- Sample size automatically embodied in posterior



Design > Condition > **Evaluate**

- Bayesian inference: Logical answer to a question in the form of a model

  "How plausible is each proportion of water, given these data?"

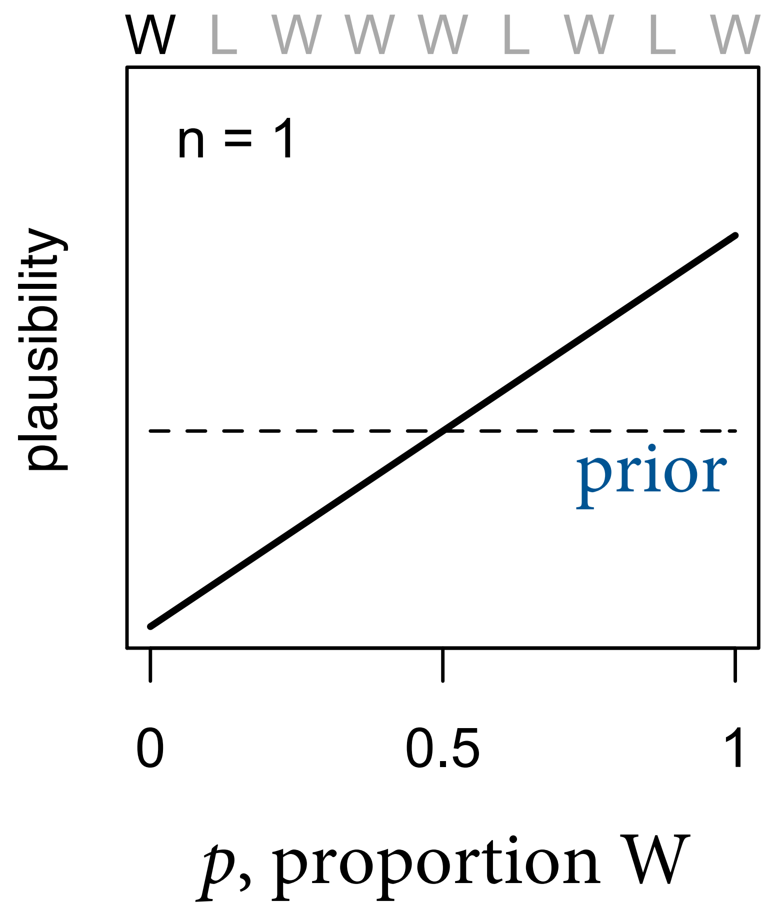
- Golem must be supervised
  - Did the golem malfunction?
  - Does the golem's answer make sense?
  - Does the question make sense?
  - Check sensitivity of answer to changes in assumptions

Definition of *W*

- Relative number of ways to see W, given N and p?
- Goal: Mathematical function to answer this question.
- The answer is a *probability distribution*.

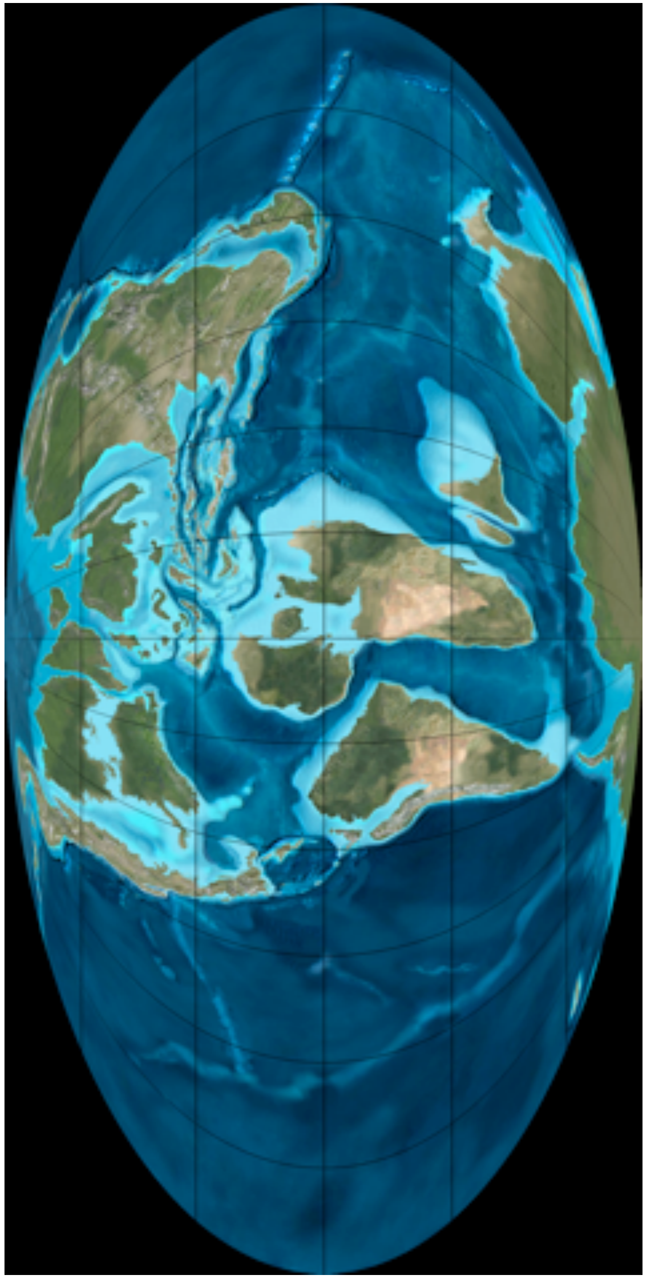
Prior probability p

- What the golem believes before the data arrive
- In this case, equal prior probability 0–1
- Pr(W) & Pr(p) define *prior predictive distribution*
- More on this later – it helps us build priors that make sense



Prior literature

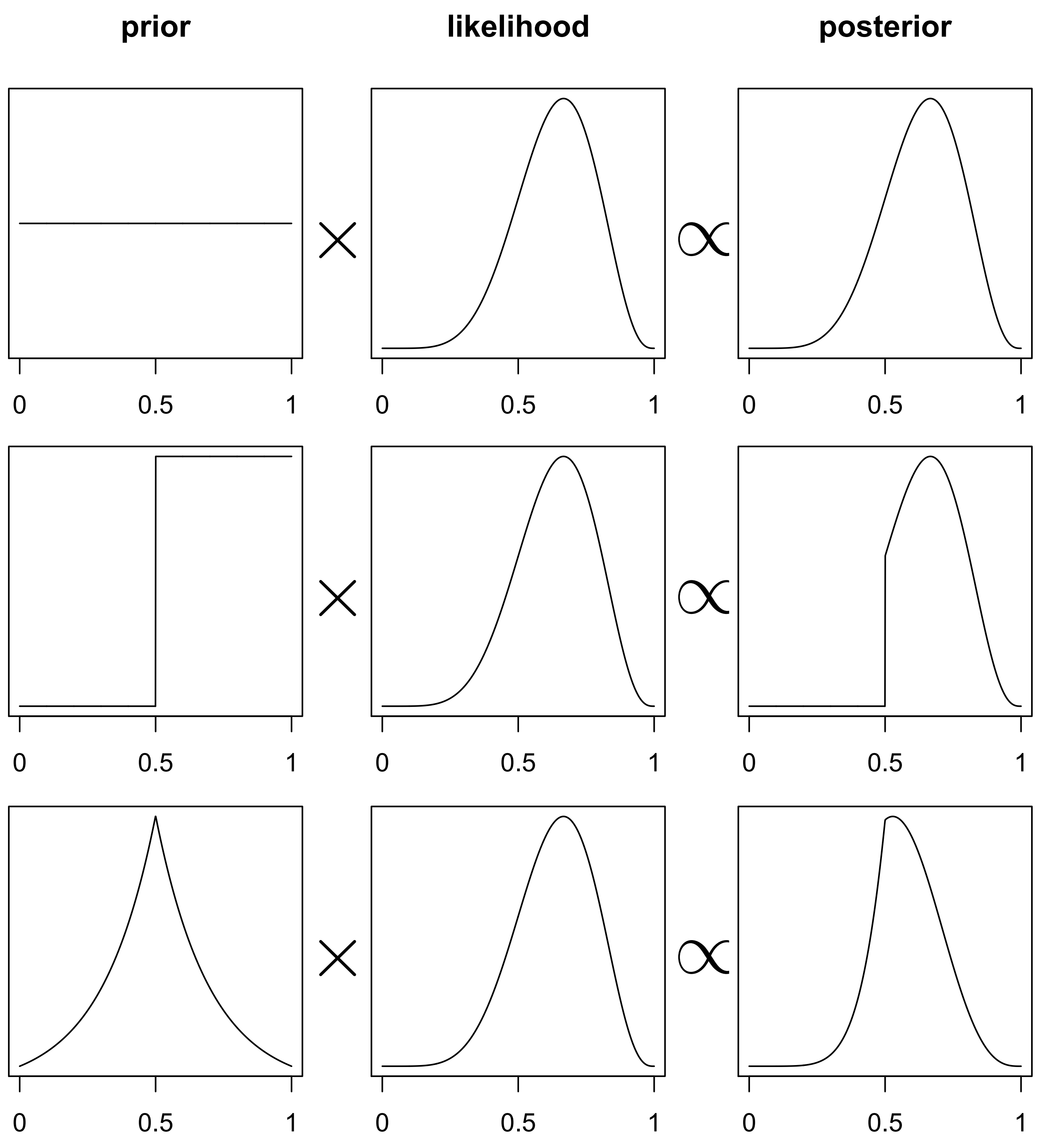
- Huge literature on choice of prior
- Flat prior conventional & bad
  - Always know something (before data) that can improve inference
  - Are zero and one plausible values for p? Is $p < 0.5$ as plausible as $p > 0.5$?
  - There is no “true” prior
  - Just need to do better than flat
- All above equally true of likelihood



Late Cretaceous (90Mya)

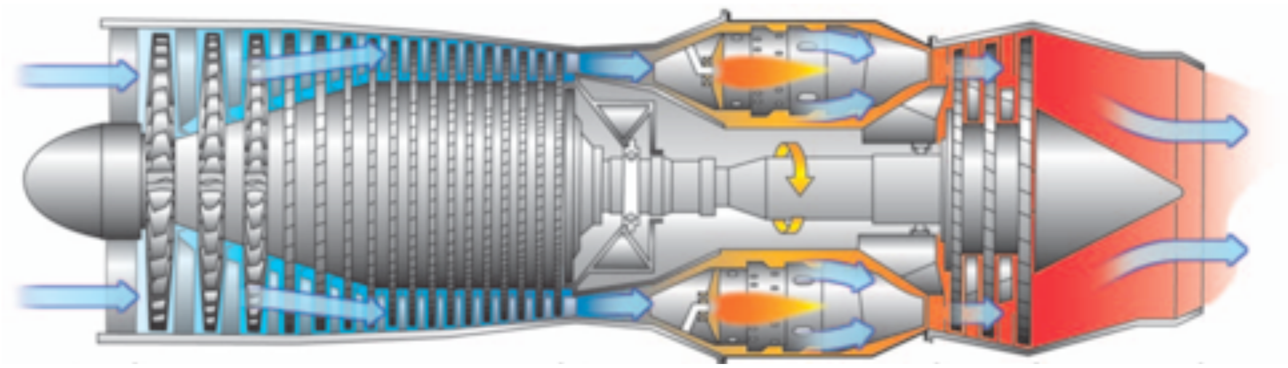
The Joint Model

$$
\begin{aligned}
W &\sim \text{Binomial}(N, p) \\
p &\sim \text{Uniform}(0, 1)
\end{aligned}
$$


prior
likelihood
posterior
0
0.5
1

Computing the posterior

1. Analytical approach (often impossible)
2. Grid approximation (very intensive)
3. Quadratic approximation (limited)
4. Markov chain Monte Carlo (intensive)



Sampling from the posterior

- Incredibly useful to *sample randomly* from the posterior
    - Visualize uncertainty
    - Compute confidence intervals
    - Simulate observations
- MCMC produces only samples
- Above all, *easier to think with samples*
- Transforms a hard calculus problem into an easy data summary problem

Sampling from the posterior

- Recipe:

  1. Compute or approximate posterior

  2. Sample with replacement from posterior

  3. Compute stuff from samples

Sample from posterior

R code 3.3

```
samples <- sample( p , prob=posterior , size=1e4 , replace=TRUE )
```

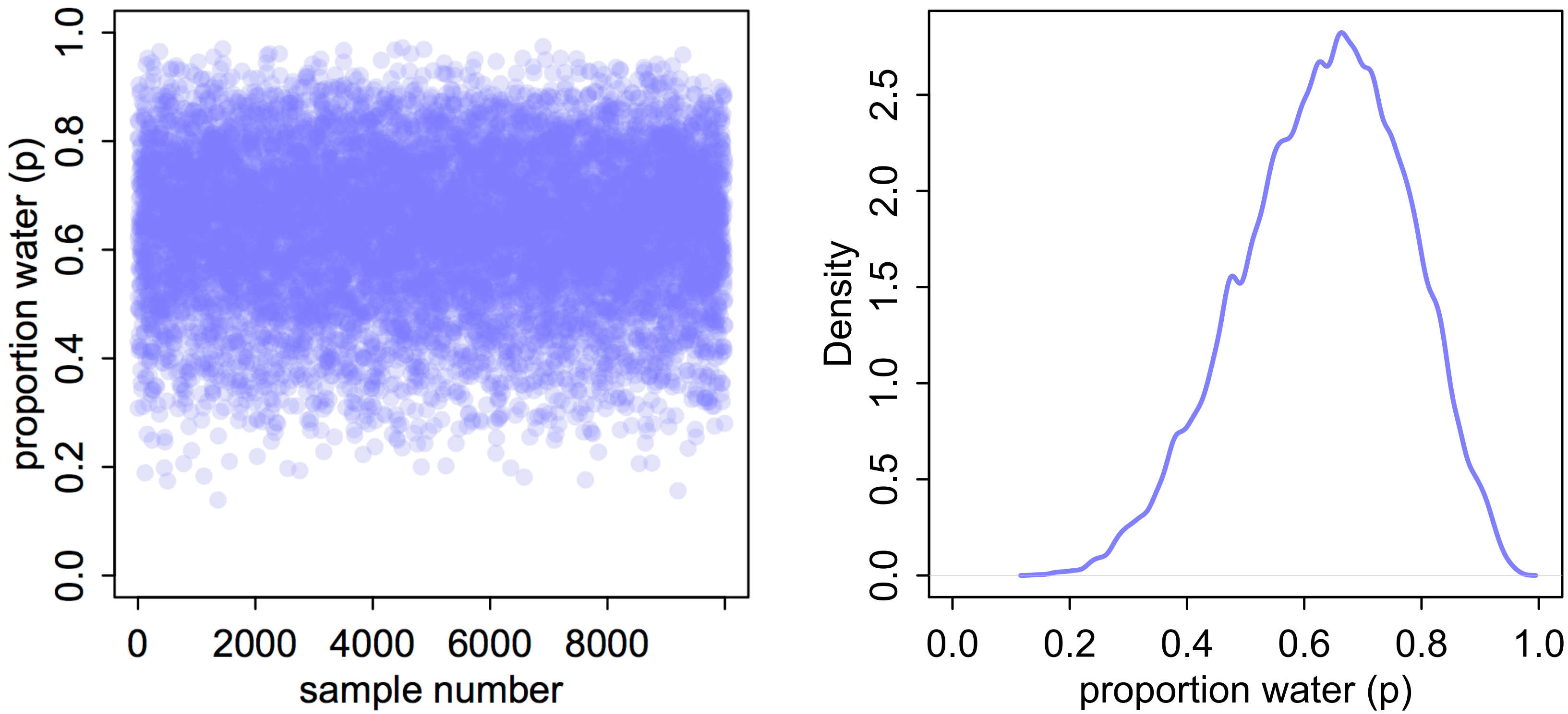


Figure 3.1

Compute stuff

- Summary tasks
  - How much posterior probability below/above/between specified parameter values?
  - Which parameter values contain 50%/80%/95% of posterior probability? "*Confidence*" *intervals*
  - Which parameter value maximizes posterior probability? Minimizes posterior loss? *Point estimates*
- You decide the question

Talking about intervals

- "Confidence interval"
  - A non-Bayesian term that doesn't even mean what it says
- "Credible interval"
  - The values are not "credible" unless you trust the model & data
- How about: *Compatibility interval*
  - Interval contains values compatible with model and data as provided
  - Small World interval

https://xkcd.com/2048/

Predictive checks

- Something like a *significance test*, but not
- No universally best way to evaluate adequacy of model-based predictions
- No way to justify always using a threshold like 5%
- Good predictive checks always depend upon purpose and imagination

"It would be very nice to have a formal apparatus that gives us some 'optimal' way of recognizing unusual phenomena and inventing new classes of hypotheses [...]; **but this remains an art for the creative human mind.**"
—E.T. Jaynes (1922–1998)